Constantly Proving The Opposite? A test of CPTO using a broad time horizon and correcting for discounting

Arthur E. Attema · Werner B. F. Brouwer

Abstract

Purpose An important assumption underlying the quality-adjusted life year (QALY) model is that people trade off life years against health in the same proportion irrespective of the number of remaining life years. This is known as the constant proportional trade-offs (CPTO) condition. Previous studies have produced mixed empirical evidence about the validity of CPTO. This paper is the first to test CPTO using the time trade-off (TTO) method for a broad time horizon.

Methods In a sample of 83 students, we use a choice based TTO protocol to elicit TTO scores for back pain, using ten different gauge durations ranging between 1 and 46 years. The TTO scores are corrected for discounting, which is elicited by means of the direct method.

Results We find average TTO scores varying between 0.72 and 0.81. Although the scores do not differ much for different durations in absolute terms, some differences are significant, rejecting CPTO, with and without correcting for discounting. No clear relationship between TTO scores and gauge duration is found. An anchoring and rounding heuristic to some extent explains our results.

Conclusions Our findings highlight the importance of elicitation methods and context dependencies in QALY measurement and warrant detailed investigation of their influence.

Keywords Time trade-off method · Constant proportional trade-off · Discounting · QALY model

A. E. Attema (✉) · W. B. F. Brouwer
iBMG/iMTA, Erasmus University Rotterdam, PO Box 1738,
3000 DR Rotterdam, The Netherlands
e-mail: attema@bmg.eur.nl

Abbreviations

CPTO	Constant proportional trade-offs
EQ-5D	EuroQol 5D
MET	Maximum endurable time
QALY	Quality-adjusted life year
TTO	Time trade-off method

Introduction

Health economic evaluations often express health benefits in terms of health-related utilities. The common utility model used in these evaluations is the quality-adjusted life years (QALY) model. The QALY model is popular mainly because it provides a straightforward way to combine the two main outcomes of health care interventions, quality of life, and life duration into a single index measure [1]. Moreover, the QALY model is intuitively appealing [2]. Despite its popularity, the QALY model suffers from severe criticism. The main objections include empirical violations of theoretical properties of the QALY model, as discussed in more detail below.

The importance of QALYs in economic evaluations renders the proper measurement of QALY tariffs to be crucial. The time trade-off (TTO) method is a widely used method to elicit these tariffs [3–5]. The purpose of the TTO method is to value health states by letting individuals trade off life years against health improvements. A common TTO exercise asks a respondent to suppose living for 10 more years in the health state to be valued, after which she dies. The respondent's task is then to indicate how many life years she is willing to give up in order to restore full health.

The properties of the QALY model translate into those of the TTO method. The descriptive deficiencies of the

QALY model cause the TTO method to be susceptible to several distorting influences, threatening its suitability to measure individuals' true preferences for health. Examples of factors leading to distortions include administration mode [6], discounting [7], loss aversion [8, 9], maximum endurable time (MET) [10, 11], and scale compatibility [12]. A result of these (and possibly other) factors in TTO is the empirical falsification of constant proportional trade-offs (CPTO, [13, 14]). This property states that health valuations obtained by the TTO method, i.e., *TTO scores*, should be the same no matter what time horizon is used in the elicitation process. In other words, individuals are assumed to trade off a constant proportion of remaining life duration in order to regain full health, regardless of the absolute size of that duration. If we, for example, would replace the 10 years in the example by 1 year, the answer of the respondent in the second task should be 1/10 of her answer in the first task. If CPTO is violated, the usual procedure to employ a 10-year time horizon in eliciting TTO scores and generalizing the results to all durations will lead to incorrect conclusions. A violation of CPTO implies that health state valuations are not constant but depend on their duration and, hence, that the QALY model is not valid. In that case, it is necessary to detect the causes of the departure from this model and to seek for modifications of the model that improve its descriptive validity.

Attema and Brouwer [14] reviewed previous studies of CPTO and showed that these studies have found negative or mixed evidence, but normally only considered rather short time horizons and, therefore, could not make an inference about the relationship between TTO scores and duration for a large number of years. The reviewed studies indicate that TTO scores tend to be high for short durations, potentially indicating that individuals do not want to give up many life years when their life expectancy is short. On the other hand, TTO scores may be higher for longer durations, because individuals may have some maximum number of life years they are prepared to sacrifice irrespective of the life expectancy in the impaired health state. Based on this reasoning, Attema and Brouwer [14] hypothesized a U shaped relationship between TTO scores and gauge duration. However, as CPTO tests had not yet been performed using a wide variety of gauge durations, no evidence existed so far that could explicitly test this hypothesis. Therefore, this paper is the first to test CPTO over a broad range of gauge durations in a within-subjects setting, both with and without controlling for discounting.

We consider a 50-year time span and elicit TTO scores for ten different durations within that time span. Moreover, we correct TTO scores for discounting and investigate the influence of this correction on the relationship between TTO scores and duration. Our procedure is entirely choice based and, hence, better embedded into economic theory

and the choice-based nature of real-life health decisions [4].

Methods

This paper uses the notation of Attema and Brouwer [14]. That is, we let $h = (h_j,\ldots, h_T)$ denote a *health profile,* where h_t represents the health state in period $t = j,\ldots, T$, with T an individual's final period of life. A *constant health profile* $h = (h_j = \alpha,\ldots, h_T = \alpha)$ is described as health profile α with duration n_α. The individual's preferences over health quality in some period are represented by the value function $v(h_t)$, while $\delta(t)$ denotes the corresponding weight attached to the value in this period. If the generalized QALY model holds, then preferences for health profiles $h = (h_j,\ldots, h_T)$ can be evaluated by the following function [15]:

$$U(t, h_t) = \sum_{t=j-1}^{T} \delta(t)v(h_t). \tag{1}$$

The term $\sum_{t=j-1}^{T} \delta(t)$ is called the *utility of life duration* for the period between $t = j - 1$ and $t = T$. In the remainder of this paper, we denote the period between two time points, e.g., x and y, $\sum_{t=x-1}^{y} \delta(t)$, by $W[x - 1, y]$. We suppress the beginning of a period if it is 0 and, hence, write $W(y)$ instead of $W[0, y]$. We normalize the utility function such that $W(0) = 0$ and $W(T) = 1$. A concave utility function for life duration is considered equivalent to *discounting* in this paper.

Ordinary CPTO holds if the proportion of remaining life years that one is willing to give up for an improvement in health status from any health state β to any health state γ does not depend on the absolute number of remaining life years [16]. If this is valid, then the utility function of life duration has to be a power function (with the linear function as a special case) [16]. This means that there exists a number $q \geq 0$, such that $q = n_\gamma/n_\beta$ and individuals are willing to give up the same proportion $(1 - q)$ of lifetime irrespective of its duration (n_β). In other words, the ratio of the number of years in γ (e.g., full health in most TTO exercises) to the number of years in β (e.g., back pain) should be the same no matter what number of years in β is chosen. One can, therefore, test ordinary CPTO by simply comparing uncorrected TTO scores, without having to know the utility of life duration function. If ordinary CPTO holds, the utility of life duration will be a member of the power family and $v(\beta)$ will have the same value irrespective of the stated period n_β.

Attema and Brouwer [14] explained that if ordinary CPTO is violated for uncorrected TTO scores, it does not necessarily follow that CPTO for corrected TTO scores

(generalized CPTO) is violated as well. *Generalized CPTO* means that the proportion of remaining *utility* of life years, $1 - W(n_\gamma)/W(n_\beta)$, that one is willing to give up for an improvement in health status from any health state β to any health state γ does not depend on the *utility* of the absolute number of remaining life years, $W(n_\beta)$ [14]. Then, there exists a number $q \geq 0$ such that $q = W(n_\gamma)/W(n_\beta)$ and individuals are willing to give up the same proportion $(1 - q)$ of utility of life duration irrespective of its duration (n_β). If generalized CPTO is also violated, this indicates a falsification of the generalized QALY model, whereas a violation of ordinary CPTO only indicates a violation of some *parametric family* of the QALY model (i.e., the power family).

If we do not make assumptions about utility of life duration curvature in TTO elicitations, we need more information before we are able to estimate health state utilities. Knowledge of the values of the durations in full health (n_{FH}) and in the impaired health state (n_β) is no longer enough, and we have to infer information about the utility function for life duration. In terms of the generalized QALY model, an ordinary TTO elicitation gives the following equation:

$$W(n_\beta)v(\beta) = W(n_{FH}). \tag{2}$$

The values of n_β and n_{FH} are known, but in order to get an estimate of $v(\beta)$, we also have to elicit $W(n_\beta)$ and $W(n_{FH})$.

These elicitations allow us to estimate q without the discounting bias, and, hence, to test the generalized CPTO property. Using this approach, the present paper tests the linear and generalized versions of the QALY model by performing a test of CPTO both for the absolute number of life years and for utilities. Some studies have tested CPTO while correcting for discounting and found both supporting [10, 16, 17] and rejecting evidence [14, 18]. However, thus far no study has tested (generalized) CPTO over a broad time horizon. Our study performs this test and addresses a number of additional questions. First, does generalized CPTO hold? If not, is there a relationship between TTO scores and gauge durations, like, for example, a U shape? Finally, what are the implications for the QALY model?

Experiment

Subjects

The subject pool consisted of 83 business administration undergraduate students who participated for course credits.

Procedure

The experiment was administered on computers in the Erasmus Behavioral Laboratory at Erasmus University Rotterdam. The experimental sessions were run by one of the authors with four subjects at a time. The subjects were separated by partitions, in order to avoid discussion between them. The sessions lasted 30 min on average.

Both the TTO part and the discounting part were choice based and used a midpoint technique to elicit indifferences [19]. Practice questions and repeat choices were included to test the understanding of the subjects. The repeat questions consisted of a repetition of the first question of a sequence at the end of that sequence. In case the choice in the repeat question disagreed with the choice in the original question, the sequence was elicited anew.

Discounting procedure

We used the risk-free utility of life duration elicitation method (*direct method*) [19] to elicit the discount weights. An advantage of this method is that it involves no uncertainty, and therefore, is not subject to distortions such as violations of expected utility. Another advantage is that the method is nonparametric, so no, possibly erroneous, parametric assumptions (e.g., exponential discounting) have to made [19]. The subjects' task in this method is to compare two different health profiles, Profile I and Profile II, each consisting of two health states: B and G, with G strictly better than B. In Profile I, the subject gets an immediate improvement in health from B to G, which lasts until time point m, after which the subject returns to health state B until point T: Profile I $= (G_1,\ldots, G_m, B_{m+1},\ldots, B_T)$. In Profile II, he starts in health state B and will be in that health state until time point m, followed by the health improvement toward health state G, which lasts until time point T: Profile II $= (B_1,\ldots, B_m, G_{m+1},\ldots, G_T)$. Let us give an example of the implementation of the direct method by describing the first question. In the first question, we set $m = 25$, equal to half the value of T (which was set at 50 years, a plausible amount for our sample of students [average age 20.5 years, SD 2.8]). Profile I was then given by $(G_1,\ldots, G_{25}, B_{26},\ldots, B_{50})$ and Profile II by $(B_1,\ldots, B_{25}, G_{26},\ldots, G_{50})$. If the subject preferred I, then the value of m was lowered; whereas, it was increased if he chose II. We went on this way until the subject was about indifferent between I and II. Attema et al. [19] showed that estimates of $W(m) = 1/2\, W(T)$ can be obtained in this way and, because we normalize $W(T)$ to 1, we obtain the simplified expression $W(m) = 1/2$.

Figure 1 illustrates the above situation in terms of the lifetime utility generated by the two profiles for $m = 10$ and $T = 20$. Profile I starts with the better health state
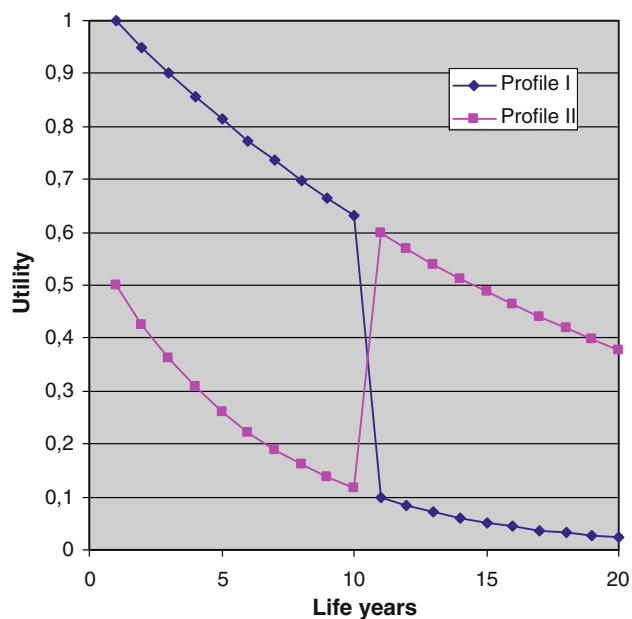
Fig. 1 Example of the lifetime utilities of two profiles

G and, hence, gives more utility at the beginning than Profile II. Because this individual discounts future life years, however, both health states are given less weight when occurring later in life. Therefore, both curves are downward sloping. Then, at point $m = 10$, health in Profile I deteriorates to B, whereas health in Profile II improves to G. This is illustrated by the kinks in the curves. They then continue decreasing smoothly (determined by the individual's discounting function) from $m = 10$ until $T = 20$. The subject now in fact compares the areas under these two curves and chooses the profile that generates the greatest area or, in other words, gives the highest discounted utility.

We described the health profiles in terms of periods of relief, i.e., for each profile we indicated during which period the subject was relieved from complaints (and, hence, being in G) and during which period he was in B. In addition, we used the age the subject would have at the indicated time points, since pilot studies suggested this was easier to imagine for subjects. Figure 2 shows a screenshot of one of the experimental questions (translated into English).

A difference in the procedure employed in [19] was that we explicitly gave the amount of time associated with a particular period with relief, in order to avoid mistakes in computations by subjects. For example, when a subject had to choose between relief from age 30 until 40 and relief from age 40 until age 60, we explicitly indicated that the relief lasted 10 years in the first option and 20 years in the second option.

TTO procedure

The first choice of a TTO iteration was always between $A =$ "n_β years in β" and $B =$ "n_β years in γ". This was to test whether subjects indeed preferred living a given period in γ rather than in β. If a subject chose A, we increased the duration of B to $n_\beta + 1$ years in γ, so that this option had both a longer duration and better health state than A. If the subject still chose A, he went forward to the next iteration and his results were not analyzed (there were seven of those subjects, as described below). If he instead chose B, we again reduced the duration of B to n_β, thereby repeating the first question. If the subject chose A again, his results were not analyzed. If he chose B, he moved on to the second question (which the subjects who preferred B already got in the first question). The second option halved the duration of B to ½n_β. Choosing B then halved this duration again, to ¼n_β, whereas choosing A led to a value halfway between ½n_β and n_β, i.e. ¾n_β. The iteration continued in this way using the bisection method, and an indifference value was estimated after five questions. There was no separate procedure for subjects who regarded the health state as worse than dead. However, it seems that virtually no subjects had this preference since for all durations there were no or very

Fig. 2 Screenshot of a question in the discounting task of the experiment

few occurrences of the lowest possible value, and the health state was explicitly chosen in such a way as to avoid worse than dead ratings.

Stimuli

We chose the constant health profiles $\beta = B$ = "regular back pain" and $\gamma = G$ = "full health" throughout the experiment. The health state "regular back pain" is a common health state, and subjects were likely to know people suffering from it. We described the health state using the domains contained in the EuroQol 5D (EQ-5D) questionnaire. We therefore indicated what regular back pain meant for daily functioning in terms of five dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression). The descriptions were printed on cards and handed to the participants (see "Appendix 1"). It was made clear to the subjects that health profile γ meant they were able to function perfectly on all five EQ-5D dimensions, irrespective of their age. The same health states were used in the discounting and TTO parts to avoid distorting influences of different behavior for different health states (e.g., if different utility functions existed for different health states).

Stimuli discounting elicitation

We elicited five points on the utility of life duration function. First, we elicited $x_{1/2} = W(m) = 1/2\ W(T = 50) = 1/2$. Subsequently, $x_{1/8}$, $x_{1/4}$, $x_{3/4}$, and $x_{7/8}$ were elicited in a similar fashion, making use of the obtained answers. For example, $x_{1/4} = W(l) = 1/2\ W(T = m) = 1/2*1/2 = 1/4$ could be elicited by setting $T = m$ and inferring a value for l in the same way as for $x_{1/2}$.

Stimuli TTO elicitation

We used ten different gauge durations in the TTO elicitation process, covering the entire range between 1 and 46 years. The gauge durations chosen were $n_\beta = 1, 3, 7, 10, 15, 19, 26, 31, 39,$ and 46 years. The use of round numbers might encourage subjects to respond in round numbers as well, causing a proportional heuristic [10]. We chose these somewhat odd durations to make this heuristic less salient [9]. Contrary to most previous studies, we used short, intermediate, and long gauge durations, enabling a more complete test of CPTO. The different gauge durations were asked in randomized order.

Analyses

We classified subjects as concave or convex depending on their five answers to the discounting elicitation questions. This procedure is explained in "Appendix 2".

The uncorrected TTO scores were computed in the usual way, i.e., dividing the elicited indifference value (number of years in full health, n_γ) by the fixed number of years with back pain (n_β). The elicited utility function for life duration was used to correct the TTO scores, employing the correcting procedure explained elsewhere [20].

We compared the results for the different gauge durations by means of the nonparametric Friedman and Wilcoxon signed ranks tests, since the TTO scores tended to be skewed to the left for all gauge durations, and a normal distribution had to be rejected (Kolmogorov–Smirnov test, $P < 0.05$ for all gauge durations). Regarding conflicting conclusions, we report parametric tests as well.

Results

The data of seven subjects were removed, because they had difficulties in understanding the experiment and did not indicate preferring more life years to less. As a result, the data of 76 subjects were included in the analysis (42 [55%] men).

Subjects always (i.e., in 100% of the questions) preferred to be in full health instead of living with back pain when the duration was the same for both, indicating that back pain was valued less than full health. Consistency tests were performed in the discounting elicitation part, yielding a test–retest reliability of 89%.

Discounting

The utility of life duration elicitation resulted mainly in concavity (or positive discounting of future life years), as expected (i.e. 244 concave parts versus 127 convex parts, binomial test: $P < 0.01$). The concavity was, however, less pronounced than in a previous elicitation by means of the direct method [19]. Classifying subjects as concave [convex] when they had at least three concave [convex] utility parts, there were 79% [21%] concave [convex] subjects in the present analysis, compared to 88% [12%] in [19]. This was probably caused by the fact that we explicitly stated the number of years corresponding to particular periods in terms of age, which may have led subjects to equalize the differences and, hence, to show more linear behavior than in the experiment of Attema et al. [19]. Figure 3 shows the utility function for the medians of the elicited life years.

TTO

Figures 4 and 5 show error bars for the TTO scores elicited with the different gauge durations, for the uncorrected and corrected TTO scores, respectively. Because the majority of the subjects exhibited positive discounting, the mean
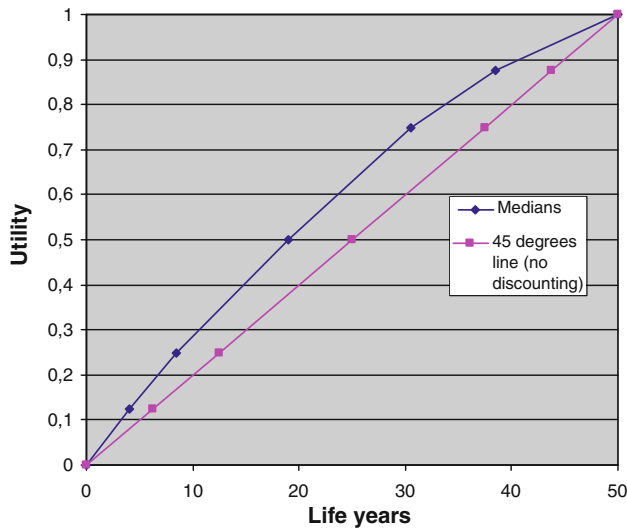
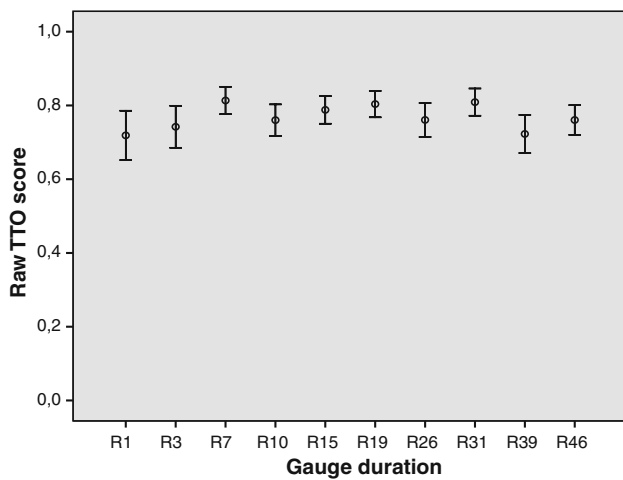Fig. 3 Utility for life duration (median *x*-values)

Fig. 4 95% confidence intervals of the estimated uncorrected TTO scores for all durations
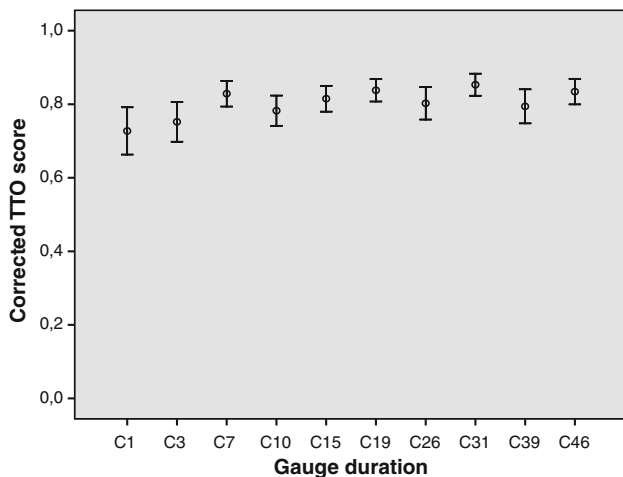

Fig. 5 95% confidence intervals of the estimated corrected TTO scores for all durations

corrected scores were higher than the mean uncorrected scores. The corrected scores were significantly higher than the uncorrected scores for all gauge durations above 1 year according to the Wilcoxon signed ranks test ($P < 0.01$ except for $n_\beta = 1$ [$P = 0.32$]) and were significantly higher for all durations above 3 years according to the paired t test ($P < 0.01$ except for $n_\beta = 1$ [$P = 0.32$] and $n_\beta = 3$ [$P = 0.06$]). Interestingly, correcting for discounting decreased the variance for all gauge durations. This seems to be an important result in light of the findings of high variability of TTO data [21].

We could not accept the hypothesis that gauge duration does not matter for TTO scores (Friedman test, $P < 0.01$), rejecting the CPTO assumption. Moreover, neither the hypothesis of a negative relationship between TTO scores and gauge duration nor the hypothesis of a U shaped relation was supported by the data. Correcting for utility of life duration decreased the magnitude of this alternation somewhat (see Fig. 5), but the conclusion did not change and the hypothesis of generalized CPTO was rejected as well (Friedman test, $P < 0.01$). There tended to be an upward trend for corrected TTO scores, i.e., lower proportions of total utility were traded off to regain full health for longer gauge durations. On the other hand, in absolute terms, the differences in mean TTO scores for the different gauge durations were fairly small. This can also be seen from Fig. 4, with most confidence intervals actually overlapping.

We also performed tests of CPTO at shorter intervals. Table 1 gives an overview of these tests, showing that CPTO could not be rejected for all adjacent gauge durations. In particular, we found no significant difference for intermediate gauge durations ($n_\beta = 15$-19-26). In addition, when using an ANOVA test for corrected scores, no significant difference for the intermediate and long durations ($n_\beta = 15$-19-26-31-39-46) was found ($P = 0.20$, although when using a Friedman test the difference was significant, $P < 0.01$). These findings provide an explanation for why some previous studies could not reject CPTO. Those studies might have considered a subset of gauge durations for which CPTO held true; whereas, it might not have been valid when they would have included a broader set of gauge durations.

Three results in particular deserve further attention. First, the 1 year gauge duration resulted in quite low TTO scores. This is contrary to the prevailing assertion that TTO scores tend to be higher for short durations, since people seem to be reluctant to give up lifetime when their life expectancy is very low. Second, there were some remarkable drops in TTO scores for the gauge durations $n_\beta = 10, 26, 39$. Third, the number of years sacrificed did not monotonically increase with gauge duration.

Table 1 P values Friedman and ANOVA tests

Test	Friedman uncorrected	Friedman corrected	ANOVA uncorrected	ANOVA corrected
Durations				
All gauge durations	<0.01	<0.01	0.02	<0.01
1–3–7	<0.01	<0.01	0.04	0.02
3–7–10	<0.01	<0.01	0.08	0.05
7–10–15	0.02	0.04	0.16	0.20
10–15–19	<0.01	<0.01	0.28	0.09
15–19–26	0.12	0.19	0.31	0.39
19–26–31	<0.01	<0.01	0.18	0.12
26–31–39	<0.01	<0.01	0.03	0.09
31–39–46	<0.01	0.02	0.02	0.08
15–19–26–31	0.02	<0.01	0.31	0.18
3–7–15–19–26–39–46	<0.01	0.02	0.04	0.05
15–19–26–31–39–46	<0.01	<0.01	0.04	0.20
15–19–26–39–46	0.01	0.15	0.08	0.42

It seems that heuristics have influenced the answers of a substantial part of the subjects. In particular, after analyzing the revealed number of life years in full health that was considered equivalent to the stated number of life years with back pain (see Table 2), we found the following peculiarities. Subjects seemed to focus on multiples of ten when making choices for the longer durations. For $n_\beta = 46$, they had a tendency to choose values around 30 and 40 years in full health. Similarly, they were inclined to choose 20 and 30 years more often for $n_\beta = 39$ and to choose values around 20 years for $n_\beta = 26$. For $n_\beta = 31$, subjects tended to take 30 as a reference point and seemed to be willing to sacrifice about 5 years, whereas many were willing to give up only 2 years for $n_\beta = 19$. These heuristics could explain the drop in TTO scores at 26 and 39 years. Furthermore, the willingness to give up no more than around 2.5 years held for $n_\beta = 10$ and $n_\beta = 15$ as well, explaining the increase in TTO scores when the gauge duration rose from 10 years toward 19 years. Thus, the absolute level of sacrificed years may have played a role here to some extent irrespective of gauge duration. Finally, for the durations shorter than 10 years, many subjects just wanted to give up the lowest possible amount, although there was more variability in the answers here causing the TTO scores not to be higher than for the longer durations.

Conclusions

Recapitulating, this study has added to the evidence against the conventional QALY model. The CPTO condition was rejected, although the magnitude of the violation was modest and the specific TTO procedure used may, because of particular heuristics, have contributed to the violations.

Correcting for discounting did not change this conclusion, so that the generalized CPTO condition was rejected as well. The correction for discounting did have other effects, though, since the TTO scores were significantly increased and variability in TTO scores decreased for all gauge durations after correcting for discounting.

Furthermore, no decreasing, increasing, U shaped, or any other clear relationship between gauge duration and TTO score was observed. We instead found an alternating pattern that was seemingly caused by anchoring heuristics. In addition, when comparing only subsets of the included gauge durations, CPTO was not always rejected, providing a possible explanation for the support for CPTO in previous empirical work. The use of long time horizons might have caused MET to become important [10, 11, 22]. However, the lack of a negative relationship between TTO scores and gauge duration suggests that the mild health state we used was not considered sufficiently serious to become worse than dead after some time.

Another, thus far neglected, phenomenon may influence TTO scores as well, i.e., the elicitation mode by which subjects reveal their indifference between two options. Many TTO studies have used some version of an open-ended, or "matching", elicitation mode to elicit TTO scores, where subjects had to give a number of years in full health that made them indifferent to the stated number of years in an imperfect health state. Instead, we used a choice task to better approximate the choice-based nature of economic practice. Our results suggest that loss aversion for short durations is far less important for choice tasks than matching tasks, in accordance with other studies [9, 23]. In particular, the shortest durations ($n_\beta = 1,3,7$) did not yield higher TTO scores than the other durations, suggesting that a choice based design causes subjects to put

Table 2 Distribution of the answers

$n_\beta = 1$		$n_\beta = 3$		$n_\beta = 7$		$n_\beta = 10$		$n_\beta = 15$	
Possible answers	Frequency	Possible answers	Frequency	Possible answers	Frequency	Possible answers	Frequency	Possible answers	Frequency
0.03	5	0.1	2	0.2	0	0.2	0	0.5	0
0.08	1	0.3	0	0.65	0	0.8	2	1.5	0
0.13	0	0.5	0	1.15	0	1.2	0	2.5	0
0.18	1	0.7	4	1.6	0	1.8	0	3.5	0
0.25	0	0.9	0	2	1	2.5	0	4.5	0
0.35	4	1.1	2	2.4	0	3.5	0	5.5	0
0.43	3	1.3	2	2.8	0	4.2	2	6.5	3
0.48	7	1.4	4	3.25	4	4.8	3	7.5	8
0.53	0	1.6	4	3.75	4	5.2	1	8.5	1
0.58	1	1.7	1	4.2	4	5.8	5	9.5	7
0.65	5	1.9	5	4.6	2	6.5	11	10.5	3
0.75	6	2.1	7	5	5	7.5	18	11.5	12
0.825	7	2.3	5	5.4	3	8.2	2	12.5	7
0.875	6	2.5	5	5.85	20	8.8	13	13.5	22
0.925	7	2.7	10	6.35	13	9.2	8	14.25	4
0.975	23	2.9	25	6.8	20	9.8	11	14.75	9

$n_\beta = 19$		$n_\beta = 26$		$n_\beta = 31$		$n_\beta = 39$		$n_\beta = 46$	
Possible answers	Frequency	Possible answers	Frequency	Possible answers	Frequency	Possible answers	Frequency	Possible answers	Frequency
0.5	0	1	3	1	0	1	2	1.5	0
1.5	0	2.5	0	3	0	3.5	0	4.5	0
3	0	3.5	0	5	0	6.5	0	7.5	0
4.5	0	5	0	7	1	9	0	10.5	0
5.5	0	7	0	9	0	11	0	13.5	1
7	1	9	0	11	0	13.5	4	16.5	2
8.5	1	11	0	13	2	16.5	3	19	2
9.5	6	12.5	4	15	3	19	13	21.5	5
10.5	2	13.5	2	17	3	21	0	24.5	3
11.5	3	15	5	19	4	23.5	0	27	2
12.5	4	17	3	21	4	26.5	3	29.5	8
13.5	3	19	16	23	3	29	13	32.5	4
14.5	7	21	8	25	10	31	1	35.5	5
15.5	7	22.5	17	27	24	33	14	38.5	17
17	30	23.5	5	29	14	35	9	41.5	20
18.5	12	25	13	30.5	8	37.5	14	44.5	7

less emphasis on the maximization of remaining lifetime. More research in this area seems warranted.

For a number of gauge durations, subjects tended to take some focal point (usually a multiple of ten) as their anchor and provided an answer close to that anchor. This anchoring heuristic offers an explanation for the alternating relationship between TTO scores and duration. Moreover, the heuristic is not a particularity of the choice design, since Attema and Brouwer [20] used a matching design and found a similar focus on 10 multiples. Other studies using longer gauge durations did not report information about individual responses, so we do not know whether these studies also found such a heuristic. Still, our findings highlight the constructive nature of health state valuation tasks, causing contextual effects to have a substantial influence on the elicited utilities. How to best avoid these heuristics is an open question.

One can question how much violation is a problem for using TTO and, as a result, the application of the QALY model in general. One may consider the variation observed in this study to be relatively small, as ranging from 0.72 to 0.81. The lack of any overall pattern combined with some computational errors may therefore be taken to suggest no strong departure from the QALY model. On the other hand, one may consider a variation of some 12.5% to be substantial. Moreover, it seems important to consider that the violation of CPTO may be more pronounced for other health states and related to the elicitation procedure used. Therefore, more research addressing these matters appears warranted.

A limitation of the current study is that we considered a student population and, hence, we could not generalize our findings to the general population. For example, age and marital status may influence discounting and health state values [24], suggesting that results may be different when interviewing older or married individuals. Moreover, short durations, i.e., life expectancies (such as of 1 year), may be difficult to realistically consider for young individuals. On the other hand, this sample did allow long time horizons to be used in our study. Another limitation may be that some students might not have been able to understand the nature of the principal health condition (i.e., back pain). Also, looking at only one disease state limited generalizability. Therefore, future studies should investigate a sample representative for the general population and include more than one health state. Finally, we used the direct method to correct for discounting, which may have particular features, e.g., that people may value descending over ascending sequences. Comparing this method to other measures of discounting is therefore important.

To conclude, our results are mixed evidence for the TTO method and the QALY model. CPTO was violated both for the usual definition and for a more generalized definition, with a variation between 0.72 and 0.81, which suggests that health quality and life duration are mutually dependent. Time preferences and heuristics are important determinants of the answers in TTO valuations, causing TTO scores to be influenced by contextual factors, like answering format (open-ended or close-ended) and availability of anchoring points. A policy implication of this study may therefore be that researchers can only compare results from using the same TTO variant, which includes anchor, elicitation procedure, and duration [25].

Appendix 1: Health state descriptions (translated)

Card 1: regular back pain

You have regular back pain. This has the following consequences for your functioning in daily life:

- You have no problems in walking about.
- You have no problems to wash or dress yourself.
- You have *some* problems with your usual activities.
- You have *moderate* pain or other discomfort.
- You are not anxious or depressed.

Card 2: full health

You have no complaints and are in full health. This has the following consequences for your functioning in daily life:

- You have no problems in walking about.
- You have no problems to wash or dress yourself.
- You have no problems with your usual activities.
- You have no pain or other discomfort.
- You are not anxious or depressed.

Appendix 2: Classification of the utility for life duration function

Classification as concave or convex was done by computing the differences between 2 successive elements of the elicited time points and dividing these by their respective utility of life duration increase (this division was necessary since the utility difference between two successive elicited values could be either 1/8 or 1/4):

$$\Delta_t = \frac{x_i - x_{i-1}}{W[x_{i-1}, x_i]} \quad i = 1, \ldots 6, \tag{3}$$

where x_1 represents $x_{1/8}$, x_i represents $x_{1/4}$, etc. Subsequently, we computed:

$$\partial_i = \Delta_i - \Delta_{i-1} \quad i = 2, \ldots 6, \tag{4}$$

i.e., how much successive outcome intervals increase or decrease per utility unit. For each subject, we observed 5 values of ∂_i. A positive [negative, zero] value of ∂_i corresponds to a concave [convex, linear] part of the utility function (i.e., positive [negative, zero] discounting) [20].

References

1. Bleichrodt, H., & Johannesson, M. (1997). The validity of QALYs: An experimental test of constant proportional tradeoff and utility independence. *Medical Decision Making, 17*(1), 21–32.

2. Miyamoto, J. M., & Eraker, S. A. (1988). A multiplicative model of the utility of survival duration and health quality. *Journal of Experimental Psychology: General, 117*(1), 3–20.

3. Dolan, P. (1997). Modeling valuations for EuroQol health states. *Medical Care, 35*(11), 1095–1108.

4. Dolan, P. (2000). The measurement of health-related quality of life for use in resource allocation decisions in health care. In A. J. Culyer & J. P. Newhouse (Eds.), *Handbook of health economics* (pp. 1723–1760). North Holland: Elsevier.

5. Burstrom, K., Johannesson, M., & Diderichsen, F. (2006). A comparison of individual and social time trade-off values for health states in the general population. *Health Policy, 76*(3), 359–370.

6. Norman, R., King, M., Clarke, D., Viney, R., Cronin, P., & Street, D. (2010). Does mode of administration matter? Comparison of online and face-to-face administration of a time trade-off task. *Quality of Life Research, 19*(4), 499–508.

7. MacKeigan, L. D., Gafni, A., & O'Brien, B. J. (2003). Double discounting of QALYs. *Health Economics, 12*(2), 165–169.

8. van Osch, S. M., Wakker, P. P., van den Hout, W. B., & Stiggelbout, A. M. (2004). Correcting biases in standard gamble and time tradeoff utilities. *Medical Decision Making, 24*(5), 511–517.

9. Bleichrodt, H., Pinto, J. L., & Abellán-Perpinán, J. M. (2003). A consistency test of the time trade-off. *Journal of Health Economics, 22*(6), 1037–1052.

10. Stalmeier, P. F., Bezembinder, T. G., & Unic, I. J. (1996). Proportional heuristics in time tradeoff and conjoint measurement. *Medical Decision Making, 16*(1), 36–44.

11. Dolan, P., & Stalmeier, P. F. M. (2003). The validity of time trade-off values in calculating QALYs: Constant proportional time trade-off versus the proportional heuristic. *Journal of Health Economics, 22*(3), 445–458.

12. Bleichrodt, H. (2002). A new explanation for the difference between time trade-off utilities and standard gamble utilities. *Health Economics, 11*(5), 447–456.

13. Tsuchiya, A., & Dolan, P. (2005). The QALY model and individual preferences for health states and health profiles over time: A systematic review of the literature. *Medical Decision Making, 25*(4), 460–467.

14. Attema, A. E., & Brouwer, W. B. F. (2010). On the (not so) constant proportional tradeoff in TTO. *Quality of Life Research, 19*(4), 489–497.

15. Bleichrodt, H., & Gafni, A. (1996). Time preference, the discounted utility model and health. *Journal of Health Economics, 15*(1), 49–66.

16. Pliskin, J. S., Shepard, D., & Weinstein, M. C. (1980). Utility functions for life years and health status. *Operations Research, 28*(1), 206–224.

17. Van der Pol, M., & Roux, L. (2005). Time preference bias in time trade-off. *European Journal of Health Economics, 6*(2), 107–111.

18. Martin, A. J., Glasziou, P. P., Simes, R. J., & Lumley, T. (2000). A comparison of standard gamble, time trade-off, and adjusted time trade-off scores. *International Journal of Technology Assessment in Health Care, 16*(1), 137–147.

19. Attema, A. E., Bleichrodt, H., & Wakker P. P. (2010). Measuring discounting and QALYs more easily and reliably. Working paper. Erasmus University Rotterdam. http://people.few.eur.nl/wakker/pdf/ulifedm.pdf. Accessed 13 August 2010.

20. Attema, A. E., & Brouwer, W. B. F. (2009). The correction of TTO-scores for utility curvature using a risk-free utility elicitation method. *Journal of Health Economics, 28*(1), 234–243.

21. Arnesen, T., & Trommald, M. (2005). Are QALYs based on time trade-off comparable? A systematic review of TTO methodologies. *Health Economics, 14*(1), 39–53.

22. Stalmeier, P. F. M., Chapman, G. B., de Boer, A. G. E. M., & van Lanschot, J. J. B. (2001). A fallacy of the multiplicative QALY model for low-quality weights in students and patients judging hypothetical health states. *International Journal of Technology Assessment in Health Care, 17*(4), 488–496.

23. Unic, I., Stalmeier, P. F., Verhoef, L. C., & van Daal, W. A. (1998). Assessment of the time-tradeoff values for prophylactic mastectomy of women with a suspected genetic predisposition to breast cancer. *Medical Decision Making, 18*(3), 268–277.

24. Dolan, P., & Roberts, J. (2002). To what extent can we explain time trade-off values from other information about respondents? *Social Science and Medicine, 54*(6), 919–929.

25. NICE. (2008). Guide to the methods of technology appraisal. http://www.nice.org.uk/aboutnice/howwework/devnicetech/technologyappraisalprocessguides/guidetothemethodsoftechnologyappraisal.jsp. Accessed 13 December 2010.